NCICB Clinical Outcomes Infrastructure
Knowledge Acquisition Session Report

Session Date: September 26, 2003 **Session Time:** 11:00am

Session Topic: User Requirements for a Clinical Trials Outcomes Infrastructure

Knowledge Analyst: Jennifer Brush, ScenPro, Inc.

Session Location: Information Sciences Building, Wing 4, City of Hope National Medical Center, Durate, CA

Type of Session:

__ Interview	__ Task Analysis	__ Scenario Analysis
__ Concept Analysis	__ Observation	_X_ Structured Interview

General Topic Area

The NCI Center for BioInformatics (NCICB) is funding an effort to develop a technical solution to the problem of providing complete and reliable clinical trial outcomes data to the cancer research community. Due to the large overall scope of developing a solution to collect, manage, report, and analyze clinical trial outcomes data, the project has been divided into multiple phases. The focus of this Phase I effort is on gathering specific user data requirements and desired system functionality.

The City of Hope is the National Data Coordinating Center for the National Comprehensive Cancer Network (NCCN). The NCCN is a non-profit alliance of 19 of the world's leading cancer centers. The NCCN Outcomes Database is a program which integrates the NCCN Practice Guidelines in Oncology with a comprehensive data collection process to measure the adherence of practice to NCCN guidelines and the clinical outcomes that patients achieve at NCCN institutions. Currently, 12 of the NCCN cancer centers are participating in the Outcomes project. This outcomes database has been in place for over six years. The purpose of this session was to document Dr's Niland and Ikle's experiences and lessons learned during the development of a web-based outcomes system.

Summary Findings

Information obtained during this session includes:

- Differences in clinical and research trial outcomes
- High level recommended Outcomes system requirements and lessons learned
- Current mechanisms and challenges associated with developing a web-based outcome system
- Challenges associated with pooling and aggregating data
- Identified goals of the NCCN Outcomes Database
- Ideas and Methods for Patient De-Identification
- Summary of high-level recommended system requirements

Report Summary

This report documents information gathered during a Knowledge Acquisition session with members of the Division of Information Sciences, City of Hope. City of Hope National Medical Center is an NCI designated Comprehensive Cancer Center. The Division of Information Sciences is the centralized research support for all City of Hope clinical research. The Division encompasses two departments: Biostatistics and Biomedial Informatics. Joyce Niland, Ph.D., is the Chair of Information Sciences and

the Director of Biostatistics. Dr. David Ikle is an Associate Director for Biostatistics.

Types of Outcomes

Clinical outcomes refers to the impact of standard care on cancer patients. Clinical outcomes data includes information about the general outcomes of how patients are responding to standard care, what treatment they received and what their measurable response was. The NCCN Outcomes Database is a clinical outcomes capture and reporting system.

Research outcomes refers to the results and effects of new therapies on patients who are enrolled in a clinical research trial. Research outcomes data includes all of the data associated with a patient as it's described in the protocol on case report forms. This includes detailed data such as lab results, prior history, prior treatment, treatment description, follow up, response assessment and long term follow up.

Outcomes Infrastructure Recommendations

Following are summaries of each of the areas Drs. Niland and Ikle recommended the NCICB take into consideration when developing a Clinical Outcomes Infrastructure. Recommendations are based on seven years experience developing the infrastructure, refining requirements, solving issues and meeting new challenges as the NCCN Outcomes database evolved. Figure 1 is a high-level representation of the NCCN Outcomes database architecture.

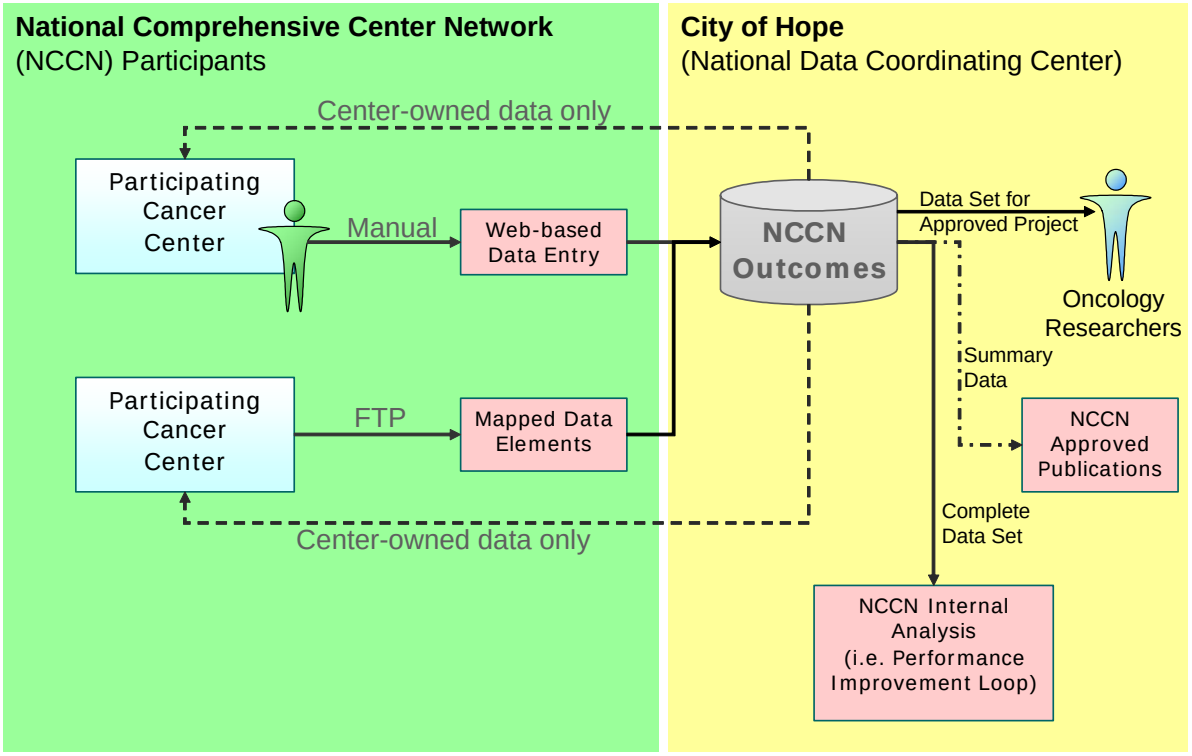

Figure 1 - NCCN Outcomes Infrastructure Notional Architecture

Data Dictionary and Standards

One of the key pieces of an outcomes data collection and reporting system is the development of a data dictionary. Shareholders must come to agreement on what information they are going to share, how it will be collected, how to code it, what standards will be followed and how the system may be

extended to include new versions of standards as they emerge. For example, the NCCN Outcomes Database is currently looking at incorporating SNOMED, as it will be available free to academic centers beginning in January of 2004.

Data Collection Options

Cancer centers may already collect the data desired for the outcomes system. Dr. Niland's approach was to give cancer centers more than one option for entering and uploading data to the NCCN's Outcomes Database. The NCCN system is web based. Users have the option of using web-based forms to manually enter data, or they may choose to upload their data via FTP.

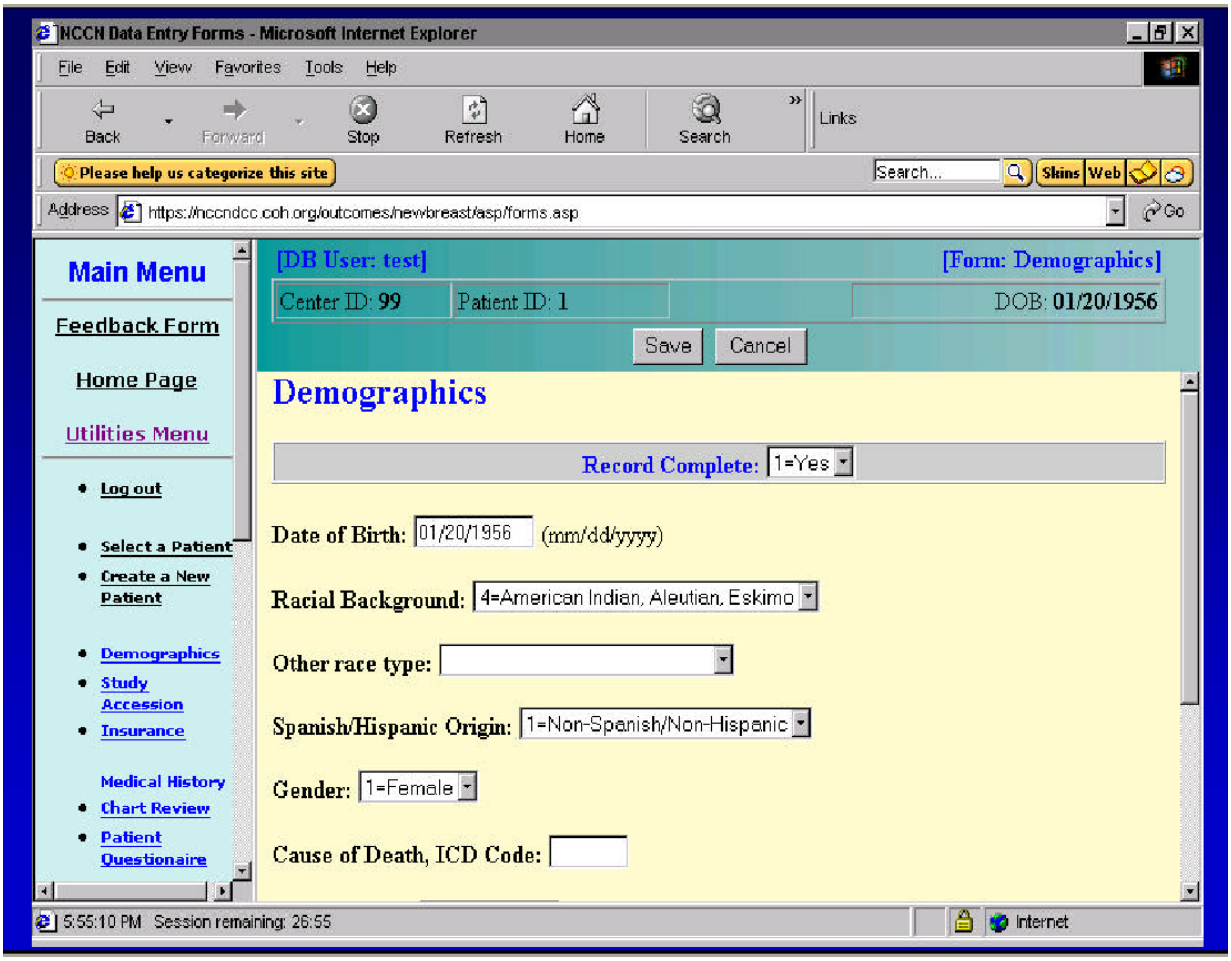

Figure 2 - Example NCCN Outcomes Data Entry Screen

The FTP option requires more initial time investment, as a participant's existing data objects have to be mapped to the NCCN Outcomes Data Dictionary. This group is also in the process of adding XML schemas to support uploading outcomes data to the system.

Data Reporting and Access

Users of the NCCN Outcomes Database also have the ability to retrieve their data online from the database at any time. Dr. Niland recommended this particular feature as a powerful incentive for any system that is built. It provides user with at least one reason for going through the effort of putting the data into the system.

The NCCN Outcomes Database has both automated (canned) and user-customizable reporting functions. The Executive Committee for the project determined the type of reports used for the automated reporting. The user-customizable portion is being piloted at a few sights. This prototype allows users to conduct coarse data mining. For example, users can select subsets of the population where staging has already been aggregated out of various fields, or whether or not patients received certain types of modalities or treatments.

Another important factor is limiting access to data that is not owned by the cancer center. In the case of NCCN, the consortium agreed to the idea of a central repository, as long as their data were secure. In order for a researcher to access data that is not from their own institution, they must make a request to the consortium to analyze the pool of data. This request goes to the Executive Committee for review. The committee evaluates the scientific concept and determines if the data is mature enough to support it. If the committee approves the request, the researcher works with the Division of Information Sciences in order to obtain the desired data sets. This is a common approach to multi-center collaborations where a researcher may only access their own data by default. In order to look at all of the available data, they have to apply to some central body for an approved project.

Additional Methods to Encourage Participation
In addition to providing multiple methods for entering data and the ability to retrieve reports about the data, the NCCN program provides the funding for one CRA (Clinical Research Assistant) to help with the data mapping and/or data entry for each cancer center.

Challenges
Resource availability is one of the biggest challenges. If the developed system is going to follow a central service model of entering data from the web, there have to be program resources to support data entry, as well as systems analysis or information management analysis at the user's end.

Linking Data Across Trials

Dr. Niland stated that there would be value in linking data across trials. A carefully designed Outcomes system could serve both the individual trial as well as longitudinal data mining. A significant benefit would be the ability to merge genetics information into the Outcomes system.

There are challenges associated with linking data across trials. Using standards, classification schemes, and approaching case report form development in a modular fashion should help with the ability to pan studies and look at the data. For example, the centralized clinical trials database being used at the City of Hope is being mined for longitudinal population studies. This is possible because the data has been kept in a standard format. Each new study utilizes the same data dictionary and elements and modules of case report forms. As a result, it serves both purposes: it goes forward to support clinical trials research, but is also supports historical data mining.

Pooling / Aggregating Trial Data

It is somewhat more difficult to pool data across studies. The whole idea of pooling should be examined carefully. There are various organizations talking about pooling and aggregating and combining data across trials, without documented guidelines about exactly how the analysis would proceed. If the studies had different treatment regimens, they could be compared, but it would be

impractical to pool them. Some data are more or less amenable to pooling. For example, it might be valuable to be able to pool Adverse Event types for a particular drug.

Studies can be compared to each other, if there is a large previously completed study. Rather than completing that study again, one could compare one study to an earlier study, compare the treatments between then and not have to reproduce the earlier one. However, anything to do with pooling or aggregating data across trials should be thought through and incorporated into the design of any system that is going to pool and aggregate data. It isn't just an informatics and a data dictionary development problem; it's a conceptual, statistical and methodological problem as well.

Outputs of an Outcomes Data Management System

There are a number of objectives for the NCCN Outcomes Database. One goal is to measure the patterns of care and see if they are concordant with the guidelines for care that have been established by the NCCN. The organization is about to publish their first paper with the results of this analysis. At a high level, the result is that cancer care is provided at all different levels to all different patient types. Every center follows the guidelines to a different degree. What that means is that patient preference, physician preference, physician biases and types of patient populations (i.e. elderly population) are influencing the care as much as having the guidelines in place.

The other output of an outcomes system like NCCN is the ability to identify prognostic factors for which patients did well and which didn't. For example, a brief outcomes study was conducted on 500 breast cancer patients at City of Hope. One of the main factors that came out of that analysis was that patients who received an early OT evaluation assessment to set up a treatment plan immediately post-surgery had improved patient response.

Outcomes data may also be used as a performance improvement loop to identify physicians who are non-concordant with the treatment guideline. For example, a center can identify a group of women who are not treated according to the guideline. Why was that the case? Was it appropriate? Sometimes there are special circumstances and patient preference. But if not, if it's just one physician's bias and he doesn't believe all the clinical trials evidence, maybe that needs to be examined more closely.

Dr. Niland's group is also looking at demographic factors in the outcomes data including: race, socioeconomic status, what level of insurance are they on, if and how that influences the care a patient receives and does it influence how well a patient responds.

Addressing Patient De-Identification

The NCCN Outcomes Database contains a limited data set. Each center created a unique participant id that is determined and maintained by that center. The centers maintain a log at their site as to what unique id they've assigned to each patient. Dr. Niland's group takes that unique id and combines it with the patient's birthday to make a unique pair. Because the data contains dates, it is not fully de-identified; it is a limited data set. The dates are retained in order to support data analysis.

Recommendations Summary

- Focus initial resources on gaining consensus about a system's data dictionary.
- Develop the system using standards, classification schemes and modular form and document development to support comparing data across trials.
- Thoroughly document standards, data aggregation and pooling schemes to support accurate data analysis.
- Provide users with multiple options for entering data into the system.
- Provide users with the ability to retrieve their own data from the system, easily, in a user-customizable format.
- Provide resources (i.e. person or funding) to a participating organization as motivation for participation.
- Conduct pilot-site studies before releasing the system to the user community at large.
- Consider providing on-site support (i.e. help desk).
- Provide centralized support for future upgrades and migrations.
- Make sure system documentation is complete and correct.